06-09-00

HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

UTILITY PATENT APPLICATION TRANSMITTAL (Only for new nonprovisional applications under 37 C.F.R. 1.53(b))		
	Attorney Docket No.	3028.1000-000
	First Named Inventor or Application Identifier	Guy A. Rouleau
	Express Mail Label No.	EL564265605US

JC530 U.S. PTO 09/590211 06/08/00

Title of Invention	SHORT GCG EXPANSIONS IN THE PAB II GENE FOR OCULOPHARYNGEAL MUSCULAR DYSTROPHY AND DIAGNOSTIC THEREOF

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

ADDRESS TO: **Assistant Commissioner for Patents, Box Patent Application, Washington, D.C. 20231**

1. [ ] Fee Transmittal Form
(Submit an original, and a duplicate for fee processing)

2. [X] Specification **Total Pages 20**
(preferred arrangement set forth below)
- Descriptive title of the invention
- Cross References to Related Applications
- Statement Regarding Fed sponsored R & D
- Reference to microfiche Appendix
- Background of the Invention
- Summary of the Invent on
- Brief Description of the Drawings
- Detailed Description
- Claim(s)
- Abstract of the Disclosure

3. [X] Drawing(s) (35 U.S.C. 113) **Total Sheets 8**
[ ] Formal [X] Informal

4. [ ] Oath or Declaration/POA **[Total Pages [ ]]**
a. [ ] Newly executed (original or copy)
b. [ ] Copy from a prior application (37 C.F.R. 1.63(d)) *(for continuation/divisional with Box 17 completed)* **[NOTE Box 5 below]**
i. [ ] DELETION OF INVENTOR(S) Signed statement attached deleting inventor(s) named in the prior application, see 37 C.F.R. 1.63(d)(2) and 1.33(b).

5. [ ] Incorporation By Reference *(useable if Box 4b is checked)* The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.

6. [ ] Microfiche Computer Program *(Appendix)*

7. [X] Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)
a. [ ] Computer Readable Copy
b. [X] Paper Copy (identical to computer copy) **3 Pages (1/3-3/3)**
c. [ ] Statement verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

8. [ ] Assignment Papers (cover sheet & documents)

9. [ ] 37 C.F.R. 3.73(b) Statement [ ] Power of Attorney *(when there is an assignee)*

10. [ ] English Translation Document *(if applicable)*

11. [ ] Information Disclosure Statement (IDS)/PTO-1449 [ ] Copies of IDS Citations

12. [ ] Preliminary Amendment

13. [X] Return Receipt Postcard (MPEP 503) *(Should be specifically itemized)*

14. [ ] Small Entity Statement(s) [ ] Statement filed in prior application, status still proper and desired

15. [ ] Certified Copy of Priority Document(s) *(if foreign priority is claimed)*

16. [X] Other: Remarks Letter

17. **If a CONTINUING APPLICATION,** check appropriate box and supply the requisite information:
[X] Continuation [ ] Divisional [ ] Continuation-in-part (CIP) of prior application No.: PCT/CA98/01133
Prior application information: Examiner: Group Art Unit:

18. CORRESPONDENCE ADDRESS

NAME	David E. Brook, Esq.				
	HAMILTON, BROOK, SMITH & REYNOLDS, P.C.				
ADDRESS	Two Militia Drive				
CITY	Lexington	STATE	MA	ZIP CODE	02421-4799
COUNTRY	USA	TELEPHONE	(781) 861-6240	FAX	(781) 861-9540

Signature	David E. Brook, R.N. 22,592 for Elizabeth W. Mata	Date	6/8/00
Submitted by Typed or Printed Name	Elizabeth W. Mata	Reg. Number	38,236

Rev. September 23, 1999

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Guy A. Rouleau and Bernard Brais

Continuation Application of International

Application No.: PCT/CA98/01133

Filed: December 7, 1998

For: SHORT GCG EXPANSIONS IN THE PAB II GENE FOR OCULOPHARYNGEAL MUSCULAR DYSTROPHY AND DIAGNOSTIC THEREOF

Date: June 8, 2000

EXPRESS MAIL LABEL NO. EL564265605US

REMARKS

Box PATENT APPLICATION
Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

The referenced application is a Continuation application claiming priority to PCT International Application No. PCT/CA98/01133, filed 7 December 1998, which claimed priority to Canadian Patent No. 2,218,199.

The application has been amended as follows from the original PCT filing: a Related Applications paragraph has been added; the description of Fig. 4 has been changed to reflect the actual labeling of the figure as Figs. 4A-4E; the specification has been amended to correct a number of typographical and syntax mistakes or irregularities; information previously residing in the Background of the Invention has been transferred to the Summary of the Invention, where it properly belongs; a description of a reference cited in the International Search report (Brais *et al.*) has been added to the Description of the Prior Art in the Background of the Invention; proper units have been inserted (*e.g.*, "μ") in the description of "PAB II mutation screening and

sequencing" in the Examples; and trademarks have been appropriately identified. No new matter has been added.

The claims in the Continuation application are based on the set of claims submitted with the Reply to the First Written Opinion with regard to the original PCT filing. The claims have been amended from those in the Reply to the First Written Opinion, in that multiple dependencies have been deleted, and the gene of Claims 1 and 31 has been specified to be an "isolated" gene.

To facilitate analysis of the claims and to emphasize that no new matter has been added, representative support in the original PCT filing is identified herein for the claims in the Continuation application. All support references pertain to pagination in the original PCT filing.

Support for Claims 1-12 and 31-32 can be found, for example, in Figure 2, at page 2, lines 12-1, at page 6, line 9, through page 7, line 6, and page 7, line 33 through page 8, line 19, and page 8, lines 26-31. Additional support for Claim 1 can also be found, for example at page 1, line 15; support for Claim 2, see Figure 2 and page 5, lines 24-30; support for Claims 5 and 6, see page 8, lines 11-12; support for Claims 7 and 8, see page 6, lines 9-17.

Support for Claim 13, can be found, for example, in Claim 6, at page 2, lines 17-31; support for Claims 14-18, which depend from Claim 13, can be found, for example, at page 7, line 33 through page 8, line 31. Support for Claims 19-24 can be found, for example, in old Claim 9, and at page 2, line 32, through page 3, line 3; support for Claims 25-27, see old Claim 10; support for Claim 28-30, see page 2, line 32, through page 3, line 3; support for Claim 33-36, see at page 7, line 33 through page 8, line 31.

If the Examiner believes that a telephone call would expedite prosecution of the application, the Examiner is invited to contact Elizabeth W. Mata at (915) 845-3558. If Elizabeth W. Mata cannot be reached, the Examiner is invited to contact David E. Brook at (781) 861-6240.

Respectfully submitted,

HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

David E. Brook RN 22,592

By for Elizabeth W. Mata

Elizabeth W. Mata
Registration No. 38,236
Telephone: (915) 845-3558
Facsimile: (915) 845-3237

Lexington, Massachusetts 02421-4799
Dated: 6/8/00

Date: June 8, 2000 Express Mail Label No. EL564265605US

Inventors: Guy A. Rouleau and Bernard Brais

Attorney's Docket No.: 3028.1000-000

SHORT GCG EXPANSIONS IN THE PAB II GENE FOR OCULO-PHARYNGEAL MUSCULAR DYSTROPHY AND DIAGNOSTIC THEREOF

RELATED APPLICATION(S)

This application is a continuation claiming priority to International Application No. PCT/CA98/01133, filed December 7, 1998 (designating the U.S.), which claims priority to Canadian Patent No. 2,218,199, filed December 9, 1997, the entire teachings of both of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

Field of the Invention

The invention relates to PAB II gene, and its uses thereof for the diagnosis, prognosis and treatment of a disease related with protein accumulation in nucleus, such as oculopharyngeal muscular dystrophy.

Description of Prior Art

Autosomal dominant oculopharyngeal muscular dys-rophy(OPMD) is an adult-onset disease with a world-wide distribution. It usually presents itself in the sixth decade with progressive swallowing difficulties (dys-phagia), eye lid drooping (ptosis) and proximal limb weakness. Unique nuclear filament inclusions in skeletal muscle fibers are its pathological hallmark (Tome, F.M.S. & Fardeau, Acta Neuropath. 49, 85-87 (1980)). Using the full power of linkage analysis in eleven French Canadian families,

the oculopharyngeal muscular dystrophy gene was fine mapped on human chromosome 14 (Brais et al., 1997, Neuromuscular Disorders 7 (Suppl.1):S70-74). A region of .75 cM was thereby identified as a region containing the potential and unknown OMPD gene (Brais et al., 1997, *supra*). Unfortunately, the OMPD gene has yet to be isolated and its nucleic acid or protein sequence have yet to be cribbed.

It would be highly desirable to be provided with a tool for the diagnosis, prognosis and treatment of a disease related with polyalanine accumulation in the nucleus, such as observed in oculopharyngeal muscular dystrophy.

SUMMARY OF THE INVENTION

One aim of the present invention is to provide a tool for the diagnosis, prognosis and treatment of a disease related with polyalanine accumulation in nucleus, such as oculopharyngeal muscular dystrophy.

Herein, the poly(A) binding protein II (PAB II) gene was isolated from a 217 kb candidate interval in chromosome 14q11.A (GCG)6 repeat encoding a polyalanine tract located at the N-terminus of the protein was expanded to (GCG)8-13 in the 144 OPMD families screened. More severe pheno-types were observed in compound heterozygotes for the (GCG)9 mutation and a (GCG)7 allele found in 2% of the population, whereas homozygosity for the (GCG)7 allele leads to autosomal recessive OPMD. Thus the (GCG)7 allele is an example of a polymorphism which can act as either a modifier of a dominant phenotype or as a recessive mutation. Pathological expansions of the polyalanine tract may cause mutated PAB II oligomers to accumulate as filament inclusions in nuclei.

In accordance with the present invention there is provided a human PAB II gene containing a transcribed polymorphic GCG repeat, which comprises a sequence as set forth in Fig. 4, which includes introns and flank-ing genomic sequence.

The allelic variants of GCG repeat of the human PAB II gene are associated with a disease related with protein accumulation in the nucleus, such as polyalanine

accumulation, or with a disease related with swallowing difficulties, such as oculopharyngeal muscular dystrophy.

In accordance with the present invention there is also provided a method for the diagnosis of a dis-ease associated with protein accumulation in the nucleus, which comprises the steps of:

a) obtaining a nucleic acid sample of said patient; and

b) determining allelic variants of a GCG repeat of the human PAB II gene; thereby long allelic variants are indicative of a disease related with protein accumulation in the nucleus, such as polyalanine accumulation and oculopharyngeal muscular dystrophy.

The long allelic variants have from about 245 to about 263 bp in length.

In accordance with the present invention there is also provided a non-human mammal model for the human PAB II gene, whose germ cells and somatic cells are modified to express at least one allelic variant of the PAB II gene and wherein said allelic variant of the PAB II is being introduced into the mammal, or an ancestor of the mammal, at an embryonic stage.

In accordance with the present invention there is also provided a method for the screening of thera-peutic agents for the prevention and/or treatment of oculopharyngeal muscular dystrophy, which comprises the steps of:

a) administering the therapeutic agents to the non-human animal of the present invention or oculopharyngeal muscular dystrophy patients; and

b) evaluating the prevention and/or treatment of development of oculopharyngeal muscular dystro-phy in this animal (such as a mammal) or in patients.

In accordance with the present invention there is also provided a method to identify genes-products thereof, or part thereof, which interact with a biochemical pathway affected by the PAB II gene, which comprises the steps of:

a) designing probes and/or primers using the PAB II gene and screening oculopharyngeal muscular dystrophy patients samples with said probes and/or primers; and

b) evaluating the role of the identified gene in oculopharyngeal muscular dystrophy patients.

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A-1B illustrate the positional cloning of the PAB II gene;

Figs. 2A-2G illustrate the OPMD (GCG)n expansion sizes and sequence of the mutation site (SEQ ID NOS:1-2);

Fig. 3 illustrates the age distribution of swallowing time (st) for French Canadian OPMD carriers of the (GCG)9 mutation; and

Figs. 4A-4E illustrate the nucleotide sequence of human poly(A) binding protein II (hPAB II)(SEQ ID NO:3).

DETAILED DESCRIPTION OF THE INVENTION

In order to identify the gene mutated in OPMD, a 350 kb cosmid contig was constructed between flanking markers D14S990 and D14S1457 (Fig. 1A). Positions of the PAB II-selected cDNA clones were determined in relation to the EcoRI restriction map and the Genealogy-based Estimate of Historical Meiosis (GEHM)-derived candidate interval (Rommens, J.M. et al., in Proceedings of the third international workshop on the identification of tran-scribed sequences (eds. Hochgeschwender, U. & Gardiner, K.) 65-79 (Plenum, New York, 1994)).

The human poly(A) binding protein II gene (PAB II) is encoded by the nucleotide sequence as set forth in Fig. 4.

Twenty-five cDNAs were isolated by cDNA selec-tion from the candidate interval (Rommens, J.M. et al., in Proceedings of the third international workshop on the identification of transcribed sequences (eds. Hochgeschwender, U. & Gardiner, K.;

65-79; Plenum, New York, 1994). Three of these hybridized to a common 20 kb EcoRI restriction fragment and showed high sequence homology to the bovine poly(A) binding protein II gene(bPAB II) (Fig. 1A). The PAB II gene appeared to be a good candidate for OPMD because it mapped to the genetically defined 0.26 cM candidate interval in 14q11 (Fig. 1A), its mRNA showed a high level of expression in skeletal muscle, and the PAB II protein is exclu-sively localized to the nucleus (Krause, S. et al., Exp. Cell Res. 214, 75-82 (1994)) where it acts as a factor in mRNA polyadenylation (Whale, E., Cell 66, 759-768 (1991); Whale, E. et al., J. Biol. Chem. 268, 2937-2945 (1993); Bienroth, S. et al., EMBO J. 12, 585-594 (1993)).

A 8 kb HindIII genomic fragment containing the PAB II gene was subcloned and sequenced (6002 bp; GenBank: AF026029)(Nemeth, A. et al., Nucleic Acids Res. 23, 4034-4041 (1995)) (Fig. 1B). Genomic structure of the PAB II gene, and position of the OPMD (GCG)n expansions. Exons are numbered. Introns 1 and 6 are variably present in 60% of cDNA clones. ORF, open read-ing frame; cen, centromere and tel, telomere.

The coding sequence was based on the previously published bovine sequence (GenBank: X89969) and the sequence of 31 human cDNAs and ESTs. The gene is comùposed of 7 exons and is transcribed in the cen-qter orientation (Fig. 1B). Multiple splice variants are found in ESTs and on Northern blots (Nemeth, A. et al., Nucleic Acids Res. 23, 4034-4041 (1995)). In particular, introns 1 and 6 are present in more than 60% of clones (Fig. 1B)(Nemeth, A. et al., Nucleic Acids Res. 23, 4034-4041 (1995)). The coding and protein sequences are highly conserved between human, bovine and mouse (GenBank: U93050). 93% of the PAB II sequence was read-ily amenable to RT-PCR- or genomic-SSCP screening. No mutations were uncovered using both techniques. How-ever, a 400 bp region of exon 1 containing the start codon could not be readily amplified. This region is 80% GC rich. It includes a (GCG)6 repeat which codes for the first six alanines of a homopolymeric stretch of 10 (Fig. 2G). Nucleotide sequence of the mutated region of PAB II as well as the amino acid

sequences of the N-terminus polyalanine stretch and position of the OPMD alanine insertions is also shown in Fig. 2.

Special conditions were designed to amplify by PCR a 242 bp genomic fragment including this GCG-repeat. The (GCG)6 allele was found in 98% of French Canadian non-OPMD control chromosomes, whereas 2% of chromosomes carried a (GCG)7 polymorphism (n=86) (Brais, B. et al., Hum. Mol. Genet. 4, 429-434 (1995)).

Screening OPMD cases belonging to 144 families showed in all cases a PCR product larger by 6 to 21 bp than that found in controls (Fig. 2A). (GCG)6 normal allele (N) and the six different (GCG)n expansions observed in 144 families.

Sequencing of these fragments revealed that the increased sizes were due to expansions of the GCG repeat (Fig. 2G). Fig. 2F shows the sequence of the (GCG)9 French Canadian expansion in a heterozygous par-ent and his homozygous child. Partial sequence of exon 1 in a normal (GCG)6 control (N), a heterozygote (ht.) and a homozygote (hm.) for the (GCG)9-repeat mutation. The number of families sharing the different (GCG)n-repeats expansions is shown in Table 1.

Table 1

Number of families sharing the different dominant (GCG)n OPMD mutations

Mutations	Polyalanine†	Families
(GCG)8	12	4
(GCG)9	13	99
(GCG)10	14	19
(GCG)11	15	16
(GCG)12	16	5
(GCG)13	17	1
Total		**144**

†, 10 alanine residues in normal PAB II.

The (GCG)9 expansion shared by 70 French Cana-dian families is the most frequent mutation we observed (Table 1).The (GCG)9 expansion is quite stable, with a single doubling observed in family F151 in an estimated 598 French Canadian meioses (Fig. 2C). The doubling of the French Canadian (GCG)9 expansion is demonstrated in Family F151.

This contrasts with the unstable nature of preùviously described disease-causing triplet-repeats (Rosenberg, R.N., New Eng. J. Med. 335, 1222-1224 (1996)).

Genotyping of all the participants in the clini-cal study of French Canadian OPMD provided molecular insights into the clinical variability observed in this condition. The genotypes for both copies of the PAB II mutated region were added to an anonymous version of this clinical database of 176 (GCG)9 mutation carriers (Brais, B. et al., Hum. Mol. Genet. 4, 429-434 (1995)). Severity of the phenotype can be assessed by the swal-lowing time (st) in seconds taken to drink 80 cc of ice-cold water (Brais, B.

et al., Hum. Mol. Genet. 4, 429-434 (1995); Bouchard, J.-P. et al., Can. J. Neurol. Sci. 19, 296-297 (1992)).The late onset and progres-sive nature of the muscular dystrophy is clearly illus-trated in heterozygous carriers of the (GCG)9 mutation (bold curve in Fig. 3) when compared to the average st of control (GCG)6 homozygous participants(n=76, thinner line in Fig. 3). The bold curve represents the average OPMD st for carriers of only one copy of the (GCG)9 mutation (n=169), while the thinner line corresponds to the average st for (GCG)6 homozygous normal con-trols(n=76). The black dot corresponds to the st value for individual VIII. Roman numerals refer to individual cases shown in Figs. 2B, 2D and discussed in the text. The genotype of a homozygous (GCG)9 patient and her parents is shown in Fig. 2B. Independent segregation of the (GCG)7 allele is also shown. Of note, case V has a more severe OPMD phenotype (Fig. 2D).

Two groups of genotypically distinct OPMD cases have more severe swallowing difficulties. Individuals I, II, and III have an early-onset disease and are homozygous for the (GCG)9 expansion ($P < 10\text{-}5$) (Figs. 2B, F). Cases IV, V, VI and VII have more severe phenotypes and are compound heterozygotes for the (GCG)9 mutation and the (GCG)7 polymorphism ($P < 10\text{-}5$). In Fig. 2D the independent segregation of the two alleles is shown. Case V, who inherited the French Canadian (GCG)9 mutation and the (GCG)7 polymorphism, is more symptomatic than his brother VIII who carries the (GCG)9 mutation and a normal (GCG)6 allele (Figs. 2D and 3). The (GCG)7 polymorphism thus appears to be a modifier of severity of dominant OPMD. Further-more, the (GCG)7 allele can act as a recessive mutaùtion. This was documented in the French patient IX who inherited two copies of the (GCG)7 polymorphism and has a late-onset autosomal recessive form of OPMD (Fig. 2E). Case IX, who has a recessive form of OPMD, is shown to have inherited two copies of the (GCG)7 polymorphism.

This is the first description of short trinu-cleotide repeat expansions causing a human disease. The addition of only two GCG repeats is sufficient to cause dominant

OPMD. OPMD expansions do not share the cardi-nal features of "dynamic mutations". The GCG expansions are not only short they are also meiotically quite sta-ble. Furthermore, there is a clear cut-off between the normal and abnormal alleles, a single GCG expansion causing a recessive phenotype. The PAB II (GCG)7 allele is the first example of a relatively frequent allele which can act as either a modifier of a dominant pheno-type or as a recessive mutation. This dosage effect is reminiscent of the one observed in a homozygote for two dominant synpolydactyly mutations. In this case, the patient had more severe deformities because she inher-ited two duplications causing an expansion in the polyalanine tract of the HOXD13 protein (Akarsu, A.N. et al., Hum. Mol. Genet. 5, 945-952 (1996)). A duplica-tion causing a similar polyalanine expansion in the a subunit 1 gene of the core-binding transcription factor (CBF(1) has also been found to cause dominant cleido-cranial dysplasia (Mundlos, S. et al., Cell 89, 773-779 (1997)). The mutations in these two rare diseases are not triplet-repeats. The are duplications of "cryptic repeats" composed of mixed synonymous codons and are thought to result from unequal crossing over (Warren, S.T., Science 275, 408-409 (1997)). In the case of OPMD, slippage during replication causing a reiteration of the GCG codon is a more likely mechanism (Wells, D.R., J. Biol. Chem. 271, 2875-2878 (1996)).

Different observations converge to suggest that a gain of function of PAB II may cause the accumulation of nuclear filaments observed in OPMD (Tome, F.M.S. & Fardeau, Acta Neuropath. 49, 85-87 (1980)). PAB II is found mostly in dimeric and oligomeric forms (Nemeth, A. et al., Nucleic Acids Res. 23, 4034-4041 (1995)). It is possible that the polyalanine tract plays a role in polymerization. Polyalanine stretches have been found in many other nuclear proteins such as the HOX pro-teins, but their function is still unknown (Davies, S.W. et al., Cell 90, 537-548 (1997)). Alanine is a highly hydrophobic amino acid present in the cores of proteins. In dragline spider silk, polyalanine stretches are thought to form B-sheet structures impor-tant in ensuring the fibers' strength (Simmons, A.H. et al., Science 271, 84-87 (1996)). Polyalanine

oligomers have also been shown to be extremely resistant to chemical denaturation and enzymatic degradation (Forood, B. et al., Bioch. and Biophy. Res. Com. 211, 7-13 (1995)). One can speculate that PAB II oligomers comprised of a sufficient number of mutated molecules might accumulate in the nuclei by forming undegradable polyalanine rich macromolecules. The rate of the accuùmulation would then depend on the ratio of mutated to non-mutated protein. The more severe phenotypes observed in homozygotes for the (GCG)9 mutations and compound heterozygotes for the (GCG)9 mutation and (GCG)7 allele may correspond to the fact that in these cases PAB II oligomers are composed only of mutated proteins. The ensuing faster filament accumulation could cause accelerated cell death. The recent descrip-tion of nuclear filament inclusions in Huntington's disease, raises the possibility that "nuclear toxicity" caused by the accumulation of mutated homopolymeric domains is involved in the molecular pathophysiology of other triplet-repeat diseases (Davies, S.W. et al., Cell 90, 537-548 (1997); Scherzinger, E. et al., Cell 90, 549-558 (1997); DiFiglia, M. et al., Science 277, 1990-1993 (1997)). Future immunocytochemical and expression studies will be able to test this patho-physiological hypothesis and provide some insight into why certain muscle groups are more affected while all tissues express PAB II.

Methods

Contig and cDNA selection

The cosmid contig was constructed by standard cosmid walking techniques using a gridded chromosome 14-specific cosmid library (Evans, G.A. et al., Gene 79, 9-20 (1989)). The cDNA clones were isolated by cDNA selection as previously described (Rommens, J.M. et al., in Proceedings of the third international workshop on the identification of transcribed sequences (eds. Hochgeschwender, U. & Gardiner, K.) 65-79 (Plenum, New York, 1994)).

Cloning of the PAB II gene

Three cDNA clones corresponding to PAB II were sequenced (Sequenase, USB). Clones were verified to map to cosmids by South-ern hybridization. The 8 kb HindIII restriction frag-ment was subcloned from cosmid 166G8 into pBluescriptII (SK) (Stratagene). The clone was sequenced using prim-ers derived from the bPABII gene and human EST sequences. Sequencing of the PAB II introns was done by primer walking.

PAB II mutation screening and sequencing

All cases were diagnosed as having OPMD on clinical grounds (Brais, B. et al., Hum. Mol. Genet. 4, 429-434 (1995)). RT-PCR- and genomic SSCP analyses were done using stan-dard protocols (Lafrenière, R.G. et al., Nat. Genet. 15, 298-302 (1997)). The primers used to amplify the PAB II mutated region were: 5'-CGCAGTGCCCCGCCTTAGA-3' (SEQ ID NO:4) and 5'-ACAAGATGGCGCCGCCGCCCCGGC-3' (SEQ ID NO:5). PCR reactions were performed in a total volume of 15 µl containing: 40 ng of genomic DNA; 1.5 µg of BSA; 1 µM of each primer; 250 µM dCTP and dTTP; 25 µM dATP; 125 µM of dGTP and 125 µM of 7-deaza-dGTP (Pharmacia); 7.5% DMSO; 3.75 µCi[^{35}S]dATP, 1.5 unit of Taq DNA polymerase and 1.5 mM MgCl2 (Perkin Elmer). For non-radioactive PCR reactions the [^{35}S]dATP was replaced by 225 µM of dATP. The amplification procedure consisted of an initial denatuùration step at 95°C for five minutes, followed by 35 cycles of denaturation at 95°C for 15 s, annealing at 70°C for 30 s, elongation at 74°C for 30 s and a final elongation at 74°C for 7 min. Samples were loaded on 5% polyacrylamide denatur-ing gels. Following electrophoresis, gels were dried and autoradiographs were obtained. Sizes of the inserts were determined by comparing to a standard M13 sequence (Sequenase™, USB). Fragments used for sequencing were gel-purified. Sequencing of the mutated fragment using the Amplicycle kit™ (Perkin Elmer) was

done with the 5'-CGCAGTGCCCCGCCTTAGAGGTG-3' (SEQ ID NO:6) primer at an elongation temperature of 68°C.

Stability of (GCG)-repeat expansions

The meiotic stability of the (GCG)9-repeat was estimated based on a large French Canadian OPMD cohort. It had been previously established that a single ancestral OPMD carrier chro-mosome was introduced in the French Canadian population by three sisters in 1648. Seventy of the seventy one French Canadian OPMD families tested to date segregate a (GCG)9 expansion. However, in family F151, the affected brother and sister, despite sharing the French Canadian ancestral haplotype, carry a (GCG)12 expansion, twice the size of the ancestral (GCG)9 mutation (Fig. 2C). In this founder effect study, it is estimated that 450 (304-594) historical meioses shaped the 123 OPMD cases belonging to 42 of the 71 enrolled families. The screening of the full set of participants allowed an identification of another 148 (GCG)9 carrier chromosomes. Therefore, it is estimated that a single mutation of the (GCG)9 expansion has occurred in 598 (452-742) meioses.

Genotype-phenotype correlations

176 carriers of at least one copy of the (GCG)9 mutation were examined during the early stage of the linkage study. All were asked to swallow 80 cc of ice-cold water as rapidly as possible. Testing was stopped after 60 seconds. The swallowing time (st) was validated as a sensitive test to identify OPMD cases (Brais, B. et al., Hum. Mol. Genet. 4, 429-434 (1995); Bouchard, J.-P. et al., Can. J. Neurol. Sci. 19, 296-297 (1992)). The st values for 76 (GCG)6 homozygotes normal controls is illustrated in Fig. 3. Analyses of variance were computed by two-way ANOVA (SYSTAT package). For the (GCG)9 homozygotes their mean st value was compared to the mean value for all (GCG)9 heterozygotes aged 35-40 ($P < 10^{-5}$). For the (GCG)9 and (GCG)7 compound

heterozygotes their mean st value was compared to the mean value for all (GCG)9 heterozygotes aged 45-65 ($P < 10^{-5}$).

While the invention has been described in con-nection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any vari-ations, uses, or adaptations of the invention follow-ing, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth, and as follows in the scope of the appended claims.

CLAIMS

What is claimed is:

1. An isolated human PAB II gene comprising a polymorphic GCG repeat in exon I thereof wherein an allelic variant of said GCG repeat is indicative of a disease associated with protein accumulation in a cell nucleus, swallowing difficulty, and/or ptosis in a human patient.

2. The gene of claim 1, wherein said polymorphic GCG repeat has the sequence

$$ATG\ (GCG)_{6+n}\ GCA,$$

wherein n is selected from 1 to 7.

3. The gene of claim 2, wherein n is selected from 2 to 7, and wherein said allelic variant is associated with an increased severity of the disease.

4. The gene of claim 3, wherein a phenotype associated with said allelic variant is dominant.

5. The gene of claim 2, wherein a first allele of said GCG repeat has an n which is equal to 1.

6. The gene of claim 5, wherein a second allele of said GCG repeat has an n selected from 2 to 7, and wherein said first allele is a modulator of the severity of the phenotype associated with said second allele.

7. The gene of claim 1, wherein said human patient is homozygous for said polymorphic GCG repeat.

8. The gene of claim 1, wherein said human patient is heterozygous for said polymorphic GCG repeat.

9. A nucleic acid sequence comprising a polymorphic GCG repeat of exon I of a human PAB II gene, wherein an allelic variant of said GCG repeat in a patient's human PAB II gene is indicative of a disease associated with protein accumulation in a cell nucleus, swallowing difficulty, and/or ptosis in said human patient.

10. The nucleic acid sequence of claim 9, wherein said polymorphic GCG repeat has the sequence

$$ATG\ (GCG)_{6+n}\ GCA,$$

wherein n is selected from 1 to 7.

11. The nucleic acid sequence of claim 10, wherein n is selected from 2 to 7, and wherein said allelic variant is associated with an increased severity of said disease.

12. The nucleic acid sequence of claim 11, wherein a phenotype associated with said allelic variant is dominant.

13. A method for the diagnosis or prognosis of a disease associated with protein accumulation in a cell nucleus, and/or swallowing difficulty and/or ptosis in a human patient, which comprises:

a) obtaining a nucleic acid sample of said patient; and

b) determining allelic variants of a GCG repeat in exon I of the PAB II gene, said GCG repeat having the sequence

$$ATG\ (GCG)_{6+n}\ GCA,$$

wherein n is selected from 0 to 7, and
whereby at least one allele of said GCG repeat having an n equal to 1 to 7, is indicative of a disease related with said protein accumulation in said nucleus, and/or a swallowing difficulty and/or ptosis in said patient.

14. The method of claim 13, wherein n is selected from 2 to 7, and wherein said allelic variant is associated with an increased severity of said disease.

15. The method of claim 14, wherein said phenotype associated with said allelic variant is dominant.

16. The method of claim 13, wherein a first allele of said GCG repeat has an n which is equal to 1.

17. The method of claim 16, wherein a second allele of said GCG repeat has an n selected from 2 to 7, and wherein said first allele is a modulator of the severity of the phenotype associated with said second allele.

18. The method of claim 13, wherein said disease is oculopharyngeal muscular dystrophy.

19. A non-human transgenic animal whose germ cells and somatic cells are modified to express at least one allelic variant of a polymorphic GCG repeat in exon I of the PAB II gene, and wherein said transgenic animal shows a phenotype of a disease associated with protein accumulation in a cell nucleus, and/or a swallowing difficulty and/or ptosis.

20. The non-human transgenic animal of claim 19, wherein said polymorphic GCG repeat has the sequence

ATG $(GCG)_{6+n}$ GCA,

wherein n is selected from 1 to 7.

21. The transgenic animal of claim 19, wherein said animal is a mammal.

22. The transgenic animal of claim 19, wherein said allelic variant of the PAB II gene is a human allelic variant.

23. The transgenic animal of claim 19, having cells which display a protein accumulation in their nucleus.

24. A cell isolated from said non-human transgenic animal according to claim 19.

25. A method for screening and identifying an agent for the prevention and/or treatment of a disease associated with protein accumulation in a cell nucleus and/or a swallowing difficulty, and/or ptosis, said method comprising:

a) exposing the transgenic animal of claim 19 to said agent; and

b) evaluating the prevention and/or treatment of development of said protein accumulation in a cell nucleus and/or a swallowing difficulty, and/or ptosis in said animal exposed to said agent as compared to a control animal not having been exposed to said agent.

26. A method for screening and identifying an agent which modulates protein accumulation in the nucleus of a cell, said method comprising:

a) exposing a cell of claim 24 to said agent; and

b) evaluating said protein accumulation in said nucleus of said exposed cell as compared to a control cell not having been exposed to said agent.

27. The method of claim 25, wherein said protein accumulation is associated with oculopharyngeal muscular dystrophy.

28. A cell which has been modified to express at least one allelic variant of a human polymorphic GCG repeat of exon I of the PAB II gene, wherein said allelic variant is associated with protein accumulation in the nucleus of said cell.

29. The cell of claim 28, wherein said polymorphic GCG repeat has the sequence

$$ATG\ (GCG)_{6+n}\ GCA,$$

wherein n is selected from 1 to 7.

30. The cell of claim 28, wherein said cell is a mammalian cell.

31. An isolated human PAB II gene comprising a polymorphic GCG repeat in exon I thereof, wherein said repeat has the sequence $ATG\ (GCG)_{6+n}\ GCA$, wherein n is 0, and wherein said sequence is indicative of a non-disease phenotype associated with protein accumulation in a cell nucleus, swallowing difficulty, and/or ptosis in a human patient.

32. The human PAB II gene of claim 31, wherein said gene is as set forth in SEQ ID NO:3.

33. A method of diagnosing a disease in a human patient associated with a meiotically stable trinucleotide expansion in a coding sequence of a gene comprising:

a) obtaining a nucleic acid sample from said patient;

b) determining whether said gene comprises at least one trinucleotide expansion, wherein the determination of one trinucleotide expansion in said coding region of said gene is indicative of a disease condition in said patient.

34. The method of claim 33, wherein said trinucleotide expansion has the sequence

$ATG(GCG)_{6+n}GCA$,

wherein n is 1 to 7.

35. The method of claim 33, wherein said disease is associated with protein accumulation in a cell nucleus and/or a swallowing difficulty and/or ptosis.

36. The method of claim 33, wherein said disease is oculopharyngeal muscular dystrophy.

SHORT GCG EXPANSIONS IN THE PAB II GENE FOR OCULO-PHARYNGEAL MUSCULAR DYSTROPHY AND DIAGNOSTIC THEREOF

ABSTRACT OF THE DISCLOSURE

The present invention relates to a human PAB II gene containing transcribed polymorphic GCG repeat, which comprises a sequence as set forth in SEQ ID NO:3, which includes introns and flanking genomic sequence. The allelic variants of GCG repeat of the human PAB II gene are associated with a disease related with protein accumulation in nucleus, such as polyalanine accumula-tion, a disease related with swallowing difficulties, such as oculopharyngeal muscular dystrophy. The pres-ent invention also relates to a method for the diagno-sis of a disease with protein accumulation in nucleus, which comprises the steps of: a) obtaining a nucleic acid sample of said patient; and b) determining allelic variants of GCG repeat of the gene of claim 1, and wherein long allelic variants are indicative of a dis-ease related with protein accumulation in nucleus.

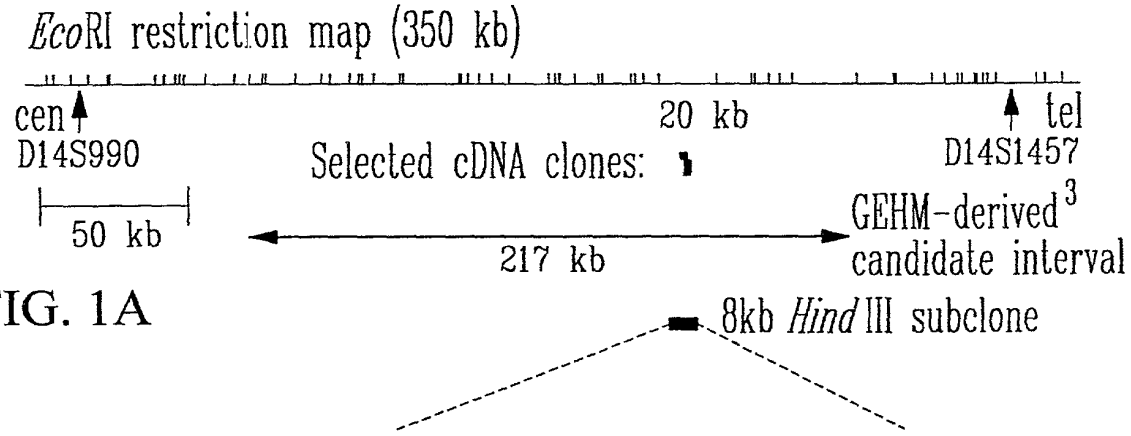


FIG. 1A

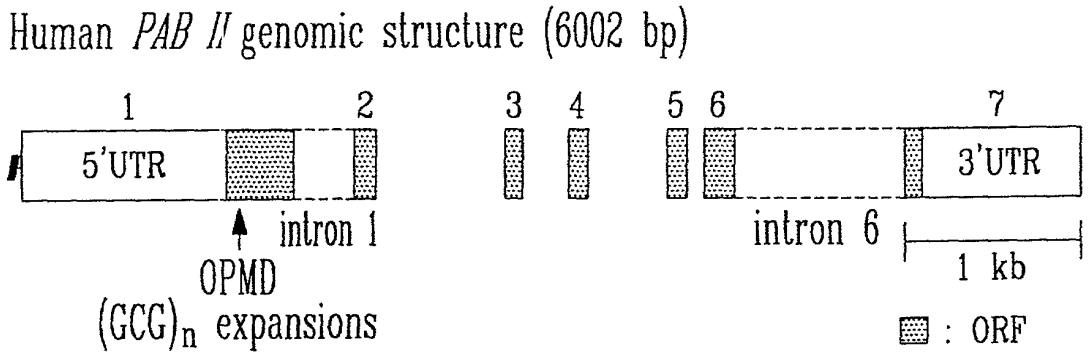


FIG. 1B

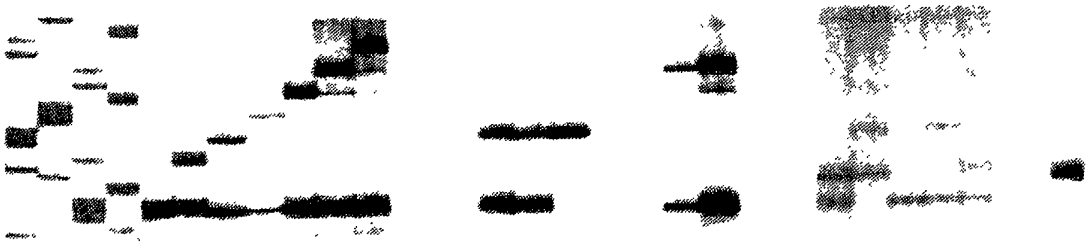
FIG. 2A

$(GCG)_{6+n}$ expansions

N2 3 4 5 6 7

FIG. 2B FIG. 2C FIG. 2D FIG. 2E

I V VIII IX



GATC

FIG. 2G

OPMD dominant mutations:

N: ATGGCGGCGGCGGCGGCGGCGGCAGCAGCA

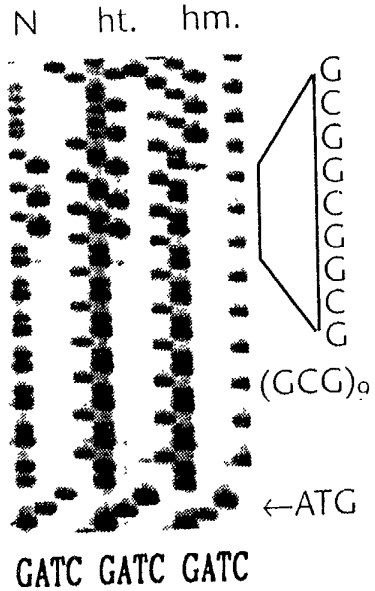
ATGGCGGCGGCGGCGGCGGCG$(GCG)_{2-7}$GCA

Polyalanine insertions:

N: MAAAAAAAAAAGAAGGRGS

MAAAAAA$(A)_{2-7}$AAAAGAAG

FIG. 2F



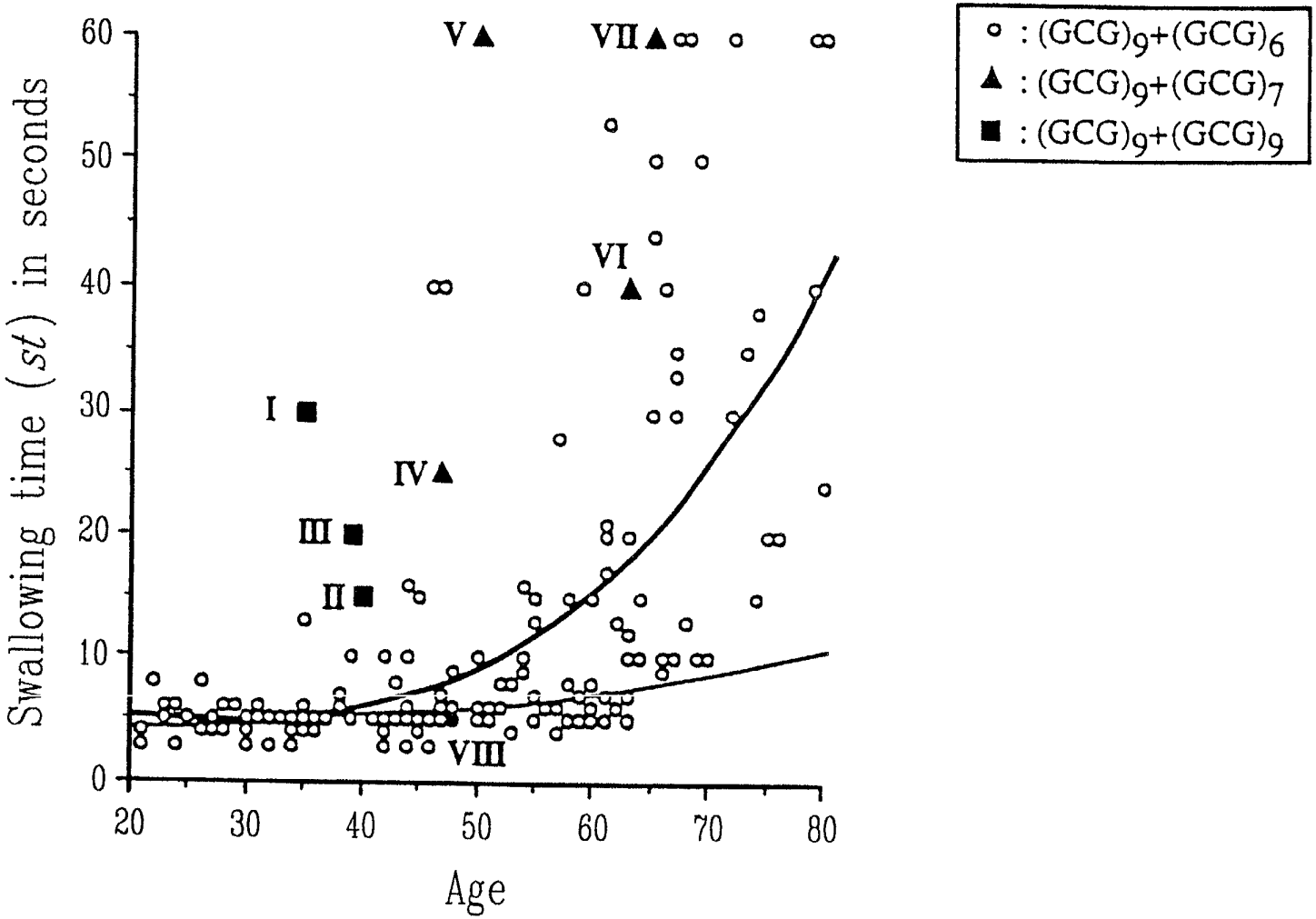


FIG. 3

```
   1 aatgaaggtg gacacccaaa tagccccaat acaaatgcct gttcaatcaa ccaaacatct
  61 aagcagcaca tctatgtggt agcatattgc caggccgtga gactgcgaat ataaatagga
 121 accgcccctc atctgcaggc gctcacaacc tagttagcaa acagtaaaac aattaagcgc
 181 gccgtggaca taggcccact tgtcctggga aatgagggga agctggggtt tgcagtggtt
 241 tgattgaagg gggactacat gttagaggca cagactgggt gcaggtacac ccaaaggaac
 301 gagaagagtg gaaggaaaca acatccacaa agtaaccaca tgctggcgta tcgaaggccg
 361 tgatttacgg ttttgagact ttacctcgcc agcaaagggg ggccagtctg ttagcggtgc
 421 agattggagg ggtgacattg gaagctgtcc aggaaaaaga aaatggaact ggggagcaga
 481 aggcctacgc aagagggcgg gacagacagg acttgtgact agtagctctg gactgaggaa
 541 tcctccctgc tttctggtgc gggagagcta gtggatgatg gtgccaataa cctggatggg
 601 gaaagtaagc tccctcctgg aatgcttcat tcacaacctc cattttcagc aacatcccat
 661 ctactggtgc ttcctggtcg agatacaagt ttcctgaaac tgctgctctg ttttgggcct
 721 cacccggcca acagctcact agctggcaag cagtagtatc aagatggcgg ccccctagga
 781 ctggctagtc atgtgacctc gggtttccca agtttgaagc ccggcagtcc tttcgggggc
 841 aaggttcacc tgtcacgaaa cgagtgtcac cccttcgact ctcgcaagcc aatcggcatc
 901 tgagactggg ccactgcggt gaggcgatcg gaagattggt cctttccagt cgcctagcta
 961 gggccaatca cggagcgtcc catacttcgc gggcccgccc gtaggccggg gagaagcagg
1021 aatatcgtca cagcgtggcg gtattattac ctaaggactc gataggaggt gggacgcgtg
1081 ttgattgaca ggcagatttc cctaccggga tttgagaatt tggcgcagtg cccgccttag
1141 aggtgcgctt atttgattgc caagtaatat tccccaatgg agtactagct catggtgacg
```

FIG. 4A

```
1201 ggcaggcagc ttgagctaat gagtcctccg tggccggcgc agctctccac atgccgggcg
1261 gcgggcccca gtctgagcgg cgatggcggc ggcggcggcg gcggcagcag cagcgggggc
1321 tgcgggcggt cggggctccg ggccggggcg gcggcgccat cttgtgcccg gggccggtgg
1381 ggaggccggg gaggggggccc cgggggggcgc aggggactac gggaacggcc tggagtctga
1441 ggaactggag cctgaggagc tgctgctgga gcccgagccg gagcccgagc ccgaagagga
1501 gccgcccccgg ccccgcgccc ccccgggagc tccgggccct gggcctggtt cgggagcccc
1561 cggcagccaa gaggaggagg aggagccggg actggtcgag ggtgacccgg gggacggcgc
1621 cattgaggac ccggtgagga aggagggcga gcgagcaggc cggcggctgg cgcgtcactg
1681 gaggcccaga gctcgggcga gcggtggcag gcgggggggtg gggttgggcg gggaataacg
1741 tggctggggc gggtcgggcc ggggatgggt cagcgatcac tacaaggggc ccgactggct
1801 tgattcgggc gtcacgggtg cctagtgttg ttctagagag ggtagctttt cttttatcac
1861 gaccctcgca tggggcgagg gaaatggccg agcatggctg aggcgcgctc tggccgagag
1921 cagggcacag cccctgcgtt ggttcctctt aagctgtcct ccataccctc cccacttata
1981 ttaggagctg gaagctatca aagctcgagt cagggagatg gaggaagaag ctgagaagct
2041 aaaggagcta cagaacgagg tagagaagca gatgaatatg agtccacctc caggcaatgc
2101 tgagtaactg gcggttgcac gcggagcccg ggttctcggg ttggaagggt tgtggggagg
2161 atggggaatg tgggqttaga tactcggcac cctggagctg cttgtctgag ctattatgac
2221 tgtgccgcgg tcatagtccg ttgtgtgttc ctctgacctt tgtgaggcag aactgatatt
2281 ttggtggtgg tagccttgtg cctccctttg tcctgttata attgtgttgc tctttattct
2341 tagtctacgt ctatctttct ttggtagagg ttgcgtgctc gcatttgacc ttcaaatcta
```

FIG. 4B

```
2401 atagtttttc ctccaattgg agacgcttta ggattctaag agaaagcaag ctggaagggg
2461 tttccccttt aaattctaga aatgtggagt ctcagcccac ttaattttgc tcactcttaa
2521 aagcatttca accaaagcca ttcattaggg atttgatttg gagggcagga gggattccta
2581 tactgtttta agtgtgtatt aattctttca atttatcgaa ttatttagtg agtaacctgc
2641 tatgcactag gcactattct cggcttgtgg gtacagcagg gaacagcaca gaccaaaatc
2701 tttgccttca ctgagcttat gggatagtgc tggtggtgga agtgcaacat attggtcaag
2761 tagaaaacaa gtgtgtggtt tttgtaaaaa attatttttt cctgatagct ggcccggtga
2821 tcatgtccat tgaggagaag atggaggctg atgcccgttc catctatgtt ggcaatgtga
2881 cgtactgggg ctctgactgg ggttgggggc aagttcttct tttggggaat tatttaatag
2941 tcctgaaaga acatctccgg gatagatgtg gttttgggtg tggagggagt gtgggaagga
3001 ggttaaaggt aatggaatga tcagtaatca gcaaaggctc tgggtttgga aggaaaagag
3061 attaattcct caaattacca gatttcatgt gctttggtgt atgatggccc agaccaaagg
3121 ctcgggaggg ttcttttgag acaggaattt gcctggtgcc tgtgaaattt ttctcctctc
3181 atcaggtgga ctatggtgca acagcagaag agctggaagc tcactttcat ggctgtggtt
3241 cagtcaaccg tgttaccata ctgtgtgaca aatttagtgg ccatcccaaa ggtaaagtaa
3301 aggggagtaa gttgagataa tttaaattac agtgtacaaa tagataaatt atgttttata
3361 ttgagcagta agttatttgg tgttaacaca ggtgatctgt gtcatttaag atcatggcat
3421 taatgttgat atatcaggag ttgcacctaa atgtcttcag aggccagata acaaaaatga
3481 aggctagatg tgggtgggat tacgaactag aaggggaggg gcagcttcta cttggcctat
3541 tatggcatat ggaaattcag gccctgtgtg tcttattttt acaaatttca aagagtagct
```

FIG. 4C

```
3601 ggaaatttta aaatttaaat gatttcgaat gattgaaatt ttccatttag aagaattttg
3661 acaaataaaa aatataactg cattgtagcc caaaacgaag catgcctgca ggttgaattt
3721 gacctgtgag gtatttgtaa cctcagagag atacaatgac aattcttttc aggtttgcgt
3781 atatagagtt ctcagacaaa gagtcagtga ggacttcctt ggccttagat gagtccctat
3841 ttagaggaag gcaaatcaag gtaagcctat gtccattgct gttctagttg tgtataaact
3901 ctccaggttg cctttaaggc tatcatttgt tcatctctga ctcaggtgat cccaaaacga
3961 accaacagac caggcatcag cacaacagac cggggttttc cacgagcccg ctaccgcgcc
4021 cggaccacca actacaacag ctcccgctct cgattctaca gtggttttaa cagcaggccc
4081 cggggtcgcg tctacaggtc aggatagatg ggctgctcct ctttcccccg cctcccgtga
4141 gccccgtatg cttcctcctc tctggtctga ggaacctccc tccccccacc cctccccgtg
4201 gtcttcagga actttgtctc ctgcctgtgc aggttgagga aggtagttgc aggccaggcc
4261 agaaggcagc ctcatcatct tttctgcagt agaaattggt gataagggct gcatccctcc
4321 cttggttcaa agaggcttcc acccccagcc ttttttttct tgggagttgg tggcatttga
4381 aggtgtttgc ggacaaaact gggaggaaca gggcctccag gaagttgaaa gcactgcttg
4441 gacatttgtt actttttttcg gagttaggga gggattgaag actgaacctc ccttggaaga
4501 ataccagagg ctagctagtt gatcctccca acagccttgt gggaggattt tgagatactt
4561 attctttatt tgagccagtc ttgcaaggtt aacttctcac tgggcctagt gtggtnccca
4621 ggtttttgcc ttgcttcact tctgtctcta catttaaata gacgggttag gcatataaac
4681 cttggctttt cataagctct acctgcctat ccccaggagt tagggaggat ctattgtga
4741 aggccctagg gtttaaaaac tgtggaggac tgaaaaactg gataaaaagg gggtcctttt
```

FIG. 4D

```
4801 ccttgcccct gtctctcact cagatgcgct tctttttcgc cactgtttgg caaagttttc
4861 tgttaagccc ccctcccccc gccccagttc tcccaggtgc gttactattt ctgggatcat
4921 ggggtcggtt ttaggacact tgaacacttc ttttcccccc ttcccttcac agtaactggg
4981 gcaggggcct acggggaggg gcttgtactg aactatctag tgatcacgtt aacacctaac
5041 tctccttctt tcttccaggg gccgggctag agcgacatca tggtattccc cttactaaaa
5101 aaagtgtgta ttaggaggag agagaggaaa aaaagaggaa agaaggaaaa aaaaaagaat
5161 taaaaaaaaa aaaaagaaaa acagaagatg accttgatgg aaaaaaaata ttttttaaaa
5221 aaaagatata ctgtggaagg ggggagaatc ccataactaa ctgctgagga gggacctgct
5281 ttggggagta ggggaaggcc cagggagtgg ggcaggggggc tgcttattca ctctggggat
5341 tcgccatgga cacgtctcaa ctgcgcaagc tgcttgccca tgtttccctg ccccccttcac
5401 ccccttgggc ctgctcaagg gtaggtgggc gtgggtggta ggagggtttt ttttacccag
5461 ggctctggaa ggacaccaaa ctgttctgct tgttaccttc cctcccgtct tctcctcgcc
5521 tttcacagtc ccctcctgcc tgctcctgtc cagccaggtc taccacccac cccacccctc
5581 tttctccggc tccctgcccc tccagattgc ctggtgatct attttgtttc cttttgtgtt
5641 tctttttctg ttttgagtgt ctttctttgc aggtttctgt agccggaaga tctccgttcc
5701 gctcccagcg gctccagtgt aaattccccc tccccctggg gaaatgcact accttgtttt
5761 gggggggttta ggggtgtttt tgtttttcag ttgttttgtt ttttttgttt ttttttttcc
5821 tttgcctttt ttccctttta tttggaggga atgggaggaa gtgggaacag ggaggtggga
5881 ggtggatttt gtttattttt ttagctcatt tccaggggtg ggaatttttt tttaatatgt
5941 gtcatgaata aagttgtttt tgaaaataaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
6001 aa
```

FIG. 4E

SEQUENCE LISTING

<110> McGILL UNIVERSITY
ROULEAU, Guy A.
BRAIS, Bernard

<120> SHORT GCG EXPANSIONS IN THE PAG II GENE
FOR OCULOPHARYNGEAL MUSCULAR DYSTROPHY AND DIAGNOSTIC
THEREOF

<130> 3028.1000-000

<150> CA 2,218,199
<151> 1997-12-09

<160> 6

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 57
<212> DNA
<213> Artificial Sequence

<400> 1
atggcggcgg cggcggccgc ggcagcagca atggcggcgg cggcggcggc ggcggca 57

<210> 2
<211> 35
<212> DNA
<213> Artificial Sequence

<400> 2
maaaaaaaaa agaaggrcsm aaaaaaaaaa agaag 35

<210> 3
<211> 6002
<212> DNA
<213> Artificial Sequence

<400> 3
aatgaaggtg gacacccaaa tagccccaat acaaatgcct gttcaatcaa ccaaacatct 60
aagcagcaca tctatgtggt agcatattgc caggccgtga gactgcgaat ataaatagga 120
accgcccctc atctgcaggc gctcacaacc tagttagcaa acagtaaaac aattaagcgc 180
gccgtggaca taggcccact tgtcctggga aatgagggga agctggggtt tgcagtggtt 240
tgattgaagg gggactacat gttagaggca cagactgggt gcaggtacac ccaaaggaac 300
gagaagagtg gaaggaaaca acatccacaa agtaaccaca tgctggcgta tcgaaggccg 360
tgatttacgg ttttgagact ttacctcgcc agcaaagggg ggccagtctg ttagcggtgc 420
agattggagg ggtgacattg gaagctgtcc aggaaaaaga aaatggaact ggggagcaga 480
aggcctacgc aagagggcgg gacagacagg acttgtgact agtagctctg gactgaggaa 540
tcctccctgc tttctggtgc gggagagcta gtggatgatg gtgccaataa cctggatggg 600
gaaagtaagc tccctcctgg aatgcttcat tcacaacctc cattttcagc aacatcccat 660
ctactggtgc ttcctggtcg agatacaagt ttcctgaaac tgctgctctg ttttgggcct 720
cacccggcca acagctcact agctggcaag cagtagtatc aagatggcgg ccccctagga 780
ctggctagtc atgtgacctc gggtttccca agtttgaagc ccggcagtcc tttcgggggc 840
aaggttcacc tgtcacgaaa cgagtgtcac cccttcgact ctcgcaagcc aatcggcatc 900
tgagactggg ccactgcggt gaggcgatcg gaagattggt cctttccagt cgcctagcta 960
gggccaatca cggagcgtcc catacttcgc gggcccgccc gtaggccggg gagaagcagg 1020

aatatcgtca cagcgtggcg gtattattac ctaaggactc gataggaggt gggacgcgtg 1080
ttgattgaca ggcagatttc cctaccggga tttgagaatt tggcgcagtg cccgccttag 1140
aggtgcgctt atttgattgc caagtaatat tccccaatgg agtactagct catggtgacg 1200
ggcaggcagc ttgagctaat gagtcctccg tggccggcgc agctctccac atgccgggcg 1260
gcgggcccca gtctgagcgg cgatggcggc ggcggcggcg gcggcagcag cagcgggggc 1320
tgcgggcggt cggggctccg ggccggggcg gcggcgccat cttgtgcccg gggccggtgg 1380
ggaggccggg gaggggggcc cgggggggcgc aggggactac gggaacggcc tggagtctga 1440
ggaactggag cctgaggagc tgctgctgga gcccgagccg gagcccgagc ccgaagagga 1500
gccgccccgg ccccgcgccc ccccgggagc tccgggccct gggcctggtt cgggagcccc 1560
cggcagccaa gaggaggagg aggagccggg actggtcgag ggtgacccgg gggacggcgc 1620
cattgaggac ccggtgagga aggagggcga gcgagcaggc cggcggctgg cgcgtcactg 1680
gaggcccaga gctcgggcga gcggtggcag gcgggggggtg gggttgggcg gggaataacg 1740
tggctggggc gggtcgggcc ggggatgggt cagcgatcac tacaaggggc ccgactggct 1800
tgattcgggc gtcacgggtg cctagtgttg ttctagagag ggtagctttt cttttatcac 1860
gaccctcgca tggggcgagg gaaatggccg agcatggctg aggcgcgctc tggccgagag 1920
cagggcacag cccctgcgtt ggttcctctt aagctgtcct ccataccctc cccacttata 1980
ttaggagctg gaagctatca aagctcgagt cagggagatg gaggaagaag ctgagaagct 2040
aaaggagcta cagaacgagg tagagaagca gatgaatatg agtccacctc caggcaatgc 2100
tgagtaactg gcggttgcac gcggagcccg ggttctcggg ttggaagggt tgtggggagg 2160
atggggaatg tggggttaga tactcggcac cctggagctg cttgtctgag ctattatgac 2220
tgtgccgcgg tcatagtccg ttgtgtgttc ctctgacctt tgtgaggcag aactgatatt 2280
ttggtggtgg tagccttgtg cctccctttg tcctgttata attgtgttgc tctttattct 2340
tagtctacgt ctatctttct ttggtagagg ttgcgtgctc gcatttgacc ttcaaatcta 2400
atagtttttc ctccaattgg agacgcttta ggattctaag agaaagcaag ctggaagggg 2460
tttccccttt aaattctaga aatgtggagt ctcagcccac ttaattttgc tcactcttaa 2520
aagcatttca accaaagcca ttcattaggg atttgatttg gagggcagga gggattccta 2580
tactgtttta agtgtgtatt aattctttca atttatcgaa ttatttagtg agtaacctgc 2640
tatgcactag gcactattct cggcttgtgg gtacagcagg gaacagcaca gaccaaaatc 2700
tttgccttca ctgagcttat gggatagtgc tggtggtgga agtgcaacat attggtcaag 2760
tagaaaacaa gtgtgtggtt tttgtaaaaa attatttttt cctgatagct ggcccggtga 2820
tcatgtccat tgaggagaag atggaggctg atgcccgttc catctatgtt ggcaatgtga 2880
cgtactgggg ctctgactcg ggttgggggc aagttcttct tttggggaat tatttaatag 2940
tcctgaaaga acatctcccg gatagatgtg gttttgggtg tggagggagt gtgggaagga 3000
ggttaaaggt aatggaatca tcagtaatca gcaaaggctc tgggtttgga aggaaaagag 3060
attaattcct caaattacca gatttcatgt gctttggtgt atgatggccc agaccaaagg 3120
ctcgggaggg ttcttttgag acaggaattt gcctggtgcc tgtgaaattt ttctcctctc 3180
atcaggtgga ctatggtgca acagcagaag agctggaagc tcactttcat ggctgtggtt 3240
cagtcaaccg tgttaccata ctgtgtgaca aatttagtgg ccatcccaaa ggtaaagtaa 3300
aggggagtaa gttgagataa tttaaattac agtgtacaaa tagataaatt atgttttata 3360
ttgagcagta agttatttgg tgttaacaca ggtgatctgt gtcatttaag atcatggcat 3420
taatgttgat atatcaggag ttgcacctaa atgtcttcag aggccagata acaaaaatga 3480
aggctagatg tgggtgggat tacgaactag aaggggaggg gcagcttcta cttggcctat 3540
tatggcatat ggaaattcag gccctgtgtg tcttattttt acaaatttca aagagtagct 3600
ggaaatttta aaatttaaat gatttcgaat gattgaaatt ttccatttag aagaattttg 3660
acaaataaaa aatataactg cattgtagcc caaaacgaag catgcctgca ggttgaattt 3720
gacctgtgag gtatttgtaa cctcagagag atacaatgac aattcttttc aggtttgcgt 3780
atatagagtt ctcagacaaa gagtcagtga ggacttcctt ggccttagat gagtccctat 3840
ttagaggaag gcaaatcaag gtaagcctat gtccattgct gttctagttg tgtataaact 3900
ctccaggttg cctttaaggc tatcatttgt tcatctctga ctcaggtgat cccaaaacga 3960
accaacagac caggcatcag cacaacagac cggggttttc cacgagcccg ctaccgcgcc 4020
cggaccacca actacaacag ctcccgctct cgattctaca gtggttttaa cagcaggccc 4080
cggggtcgcg tctacaggtc aggatagatg ggctgctcct ctttcccccg cctcccgtga 4140
gccccgtatg cttcctcctc tctggtctga ggaacctccc tccccccacc cctccccgtg 4200
gtcttcagga actttgtctc ctgcctgtgc aggttgagga aggtagttgc aggccaggcc 4260
agaaggcagc ctcatcatct tttctgcagt agaaattggt gataagggct gcatccctcc 4320
cttggttcaa agaggcttcc acccccagcc ttttttttct tgggagttgg tggcatttga 4380

aggtgtttgc ggacaaaact gggaggaaca gggcctccag gaagttgaaa gcactgcttg 4440
gacatttgtt actttttcg gagttaggga gggattgaag actgaacctc ccttggaaga 4500
ataccagagg ctagctagtt gatcctccca acagccttgt gggaggattt tgagatactt 4560
attctttatt tgagccagtc ttgcaaggtt aacttctcac tgggcctagt gtggtnccca 4620
ggtttttgcc ttgcttcact tctgtctcta catttaaata gacgggttag gcatataaac 4680
cttggctttt cataagctct acctgcctat ccccaggagt tagggaggat ctatttgtga 4740
aggccctagg gtttaaaaac tgtggaggac tgaaaaactg gataaaaagg gggtcctttt 4800
ccttgccccc gtctctcact cagatgcgct tctttttcgc cactgtttgg caaagttttc 4860
tgttaagccc ccctcccccc gccccagttc tcccaggtgc gttactattt ctgggatcat 4920
ggggtcggtt ttaggacact tgaacacttc ttttcccccc ttcccttcac agtaactggg 4980
gcaggggcct acggggaggg gcttgtactg aactatctag tgatcacgtt aacacctaac 5040
tctccttctt tcttccaggg gccgggctag agcgacatca tggtattccc cttactaaaa 5100
aaagtgtgta ttaggaggag agagaggaaa aaaagaggaa agaaggaaaa aaaaaagaat 5160
taaaaaaaaa aaaaagaaaa acagaagatg accttgatgg aaaaaaaata ttttttaaaa 5220
aaaagatata ctgtggaagg ggggagaatc ccataactaa ctgctgagga gggacctgct 5280
ttggggagta ggggaaggcc cagggagtgg ggcagggggc tgcttattca ctctggggat 5340
tcgccatgga cacgtctcaa ctgcgcaagc tgcttgccca tgtttccctg ccccccttcac 5400
ccccttgggc ctgctcaacg gtaggtgggc gtgggtggta ggagggtttt ttttacccag 5460
ggctctggaa ggacaccaaa ctgttctgct tgttaccttc cctcccgtct tctcctcgcc 5520
tttcacagtc ccctcctgcc tgctcctgtc cagccaggtc taccacccac cccacccctc 5580
tttctccggc tccctgcccc tccagattgc ctggtgatct attttgtttc cttttgtgtt 5640
tctttttctg ttttgagtct ctttctttgc aggtttctgt agccggaaga tctccgttcc 5700
gctcccagcg gctccagtct aaattccccc tccccctggg gaaatgcact accttgtttt 5760
gggggggttta ggggtgtttt tgtttttcag ttgttttgtt tttttgtttt ttttttttcc 5820
tttgcctttt ttccctttta tttggaggga atgggaggaa gtgggaacag ggaggtggga 5880
ggtggatttt gtttattttt ttagctcatt tccaggggtg ggaatttttt tttaatatgt 5940
gtcatgaata aagttgtttt tgaaaataaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 6000
aa 6002

<210> 4
<211> 19
<212> DNA
<213> Artificial Sequence

<400> 4
cgcagtgccc cgccttaga 19

<210> 5
<211> 24
<212> DNA
<213> Artificial Sequence

<400> 5
acaagatggc gccgccgccc cggc 24

<210> 6
<211> 23
<212> DNA
<213> Artificial Sequence

<400> 6
cgcagtgccc cgccttagag gtg 23